Unsupervised Classification

"Clustering"
MultiSpec© PC

A Beginning Note: The MultiSpec© software is an ongoing project of the *Laboratory for Applications of Remote Sensing*, at Purdue University. Developed by David Landgrebe and programmed by Larry Biehl, the software was originally written for the Macintosh platform. The software is being converted to the Windows platform and is now capable of carrying out all of the processing necessary for GLOBE protocols and activities. This tutorial was written for the October 1998 release of MultiSpec© PC. Users of older versions should visit the MultiSpec© web site

http://dynamo.ecn.purdue.edu/~biehl/MultiSpec

and download the newest release. All users are encouraged to visit the Purdue site periodically for information on updated versions of the software.

To use this tutorial, you will need the **Beverly.lan** image used in GLOBE training, and an image file named **beverly.clu**. These are available from the GLOBE Program web site, or they may be downloaded from the University of New Hampshire site at:

http://www.globe.unh.edu

Follow the links to "MultiSpec Tutorials."

Note: To obtain the best performance from MultiSpec© PC, you should set your monitor display to "thousands" or "millions" of colors. At 256 colors, some features of MultiSpec© may not function properly.

Unsupervised Classification
"Clustering"
MultiSpec© PC

Each pixel in your LandSat TM image contains a wealth of information about the surface materials that reflected light from that pixel to the satellite sensors. Each pixel contains a reflectance value that can range from 0 to 255 for each TM band supplied with your image. If, for example, your image contains data for five bands, then each pixel contains five pieces of data, each potentially ranging from 0 to 255, as shown in the sample pixel diagram to the right.

This means that your image could contain 256^5 (that's approximately 1.1 billion) different possible spectral combinations. Each of these combinations does not represent a different type of land cover; most of these variations represent very small and, to us, "unseeable" differences in surface conditions.

In most instances, your computer monitor will be displaying only a few thousand colors, hence a few thousand different pixels. Even set to "millions" of colors, only a small part of the many different pixels can be displayed. Even if a monitor could display all the different possible pixels, your eyes could recognize only a small number of differences in their appearance.

Because there is a limited number of different land cover types (the Modified UNESCO Classification System, MUC, contains about 157 different types), and no GLOBE study site will have more than a few of these different land cover types, it is necessary to group pixels together into a smaller number of closely related "classes." This process, whereby pixels with similar spectral characteristics are grouped, is called "Classification," and is done in two different ways.

In a supervised classification, you "train" the software to recognize that certain types of pixels represent specific land cover types. This is done on the basis of your

knowledge of your own area, and fieldwork you may do. The software then classifies the pixels of your image into the groups you have specified.

In an unsupervised classification, or "Clustering", we enter the number of groups, or "clusters," we wish to have, and certain other specifications. The software then examines the pixels in the image and groups them according to similar spectral characteristics. These groupings are not made on the basis of land cover, but on the similarity of the spectral characteristics of the pixels.

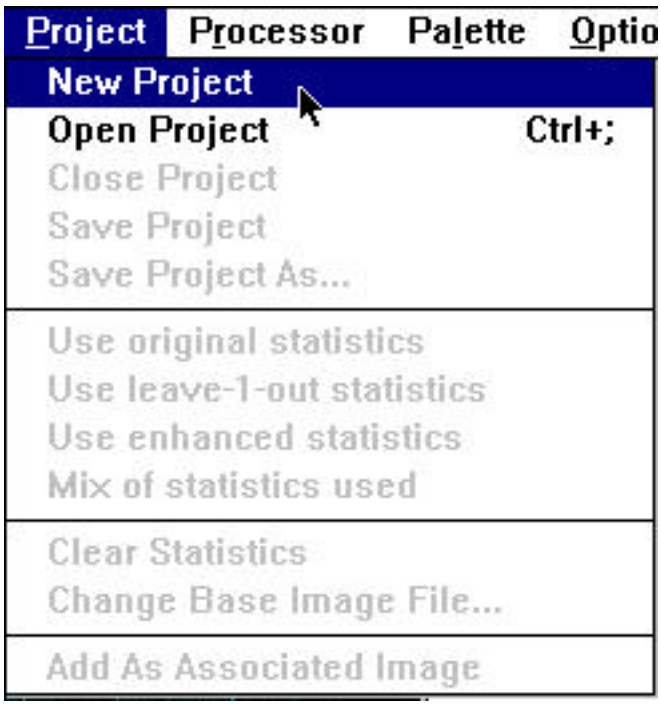
As part of your preparation of a land cover map for your 15 km x 15 km GLOBE Study Site, it is necessary for you to identify relatively large, homogeneous areas in your image for ground study and later use in validating your completed land cover map. To do this, you will have MultiSpec© cluster your image. This will help you locate areas to visit for ground verification studies.

Clustering

To demonstrate clustering, you will use the Beverly, Massachusetts (**Beverly.lan**) image provided with your MultiSpec© tutorial (See “A Beginning Note", page 1).

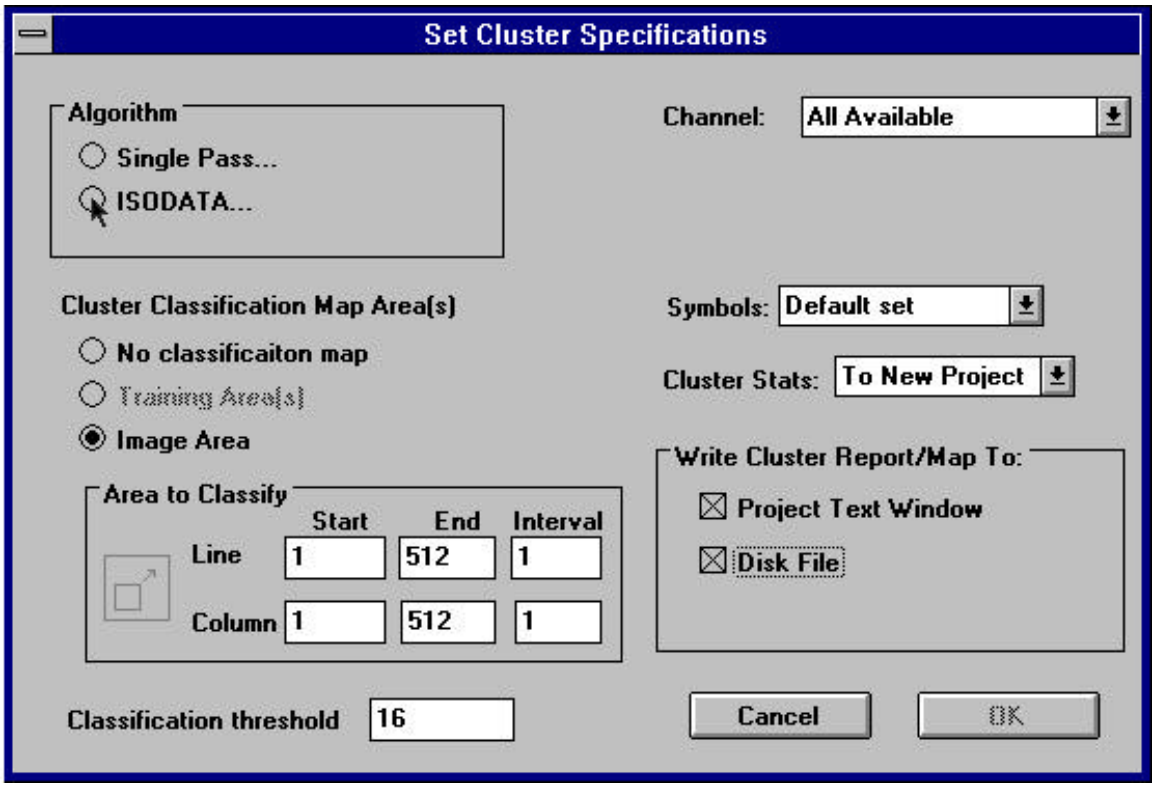
• Launch **MultiSpec©** and **Open** the **Beverly.lan** image. This is a typical 5 channel, 512 pixel x 512 pixel GLOBE Study Site image.

• From the **Project** menu, as shown below, select **New Project**.



Your clustering exercises are saved as “Projects” and, when done, can be opened by MultiSpec© as “Thematic Images.”

• From the **Processor** menu, select **Cluster**. “Clustering” is MultiSpec©’s terminology for an Unsupervised Classification. As shown below, The **Set Cluster Specifications** window opens. It is in this window that you select a clustering “algorithm" (method by which the software carries out its clustering) and enter certain values for the software to use.

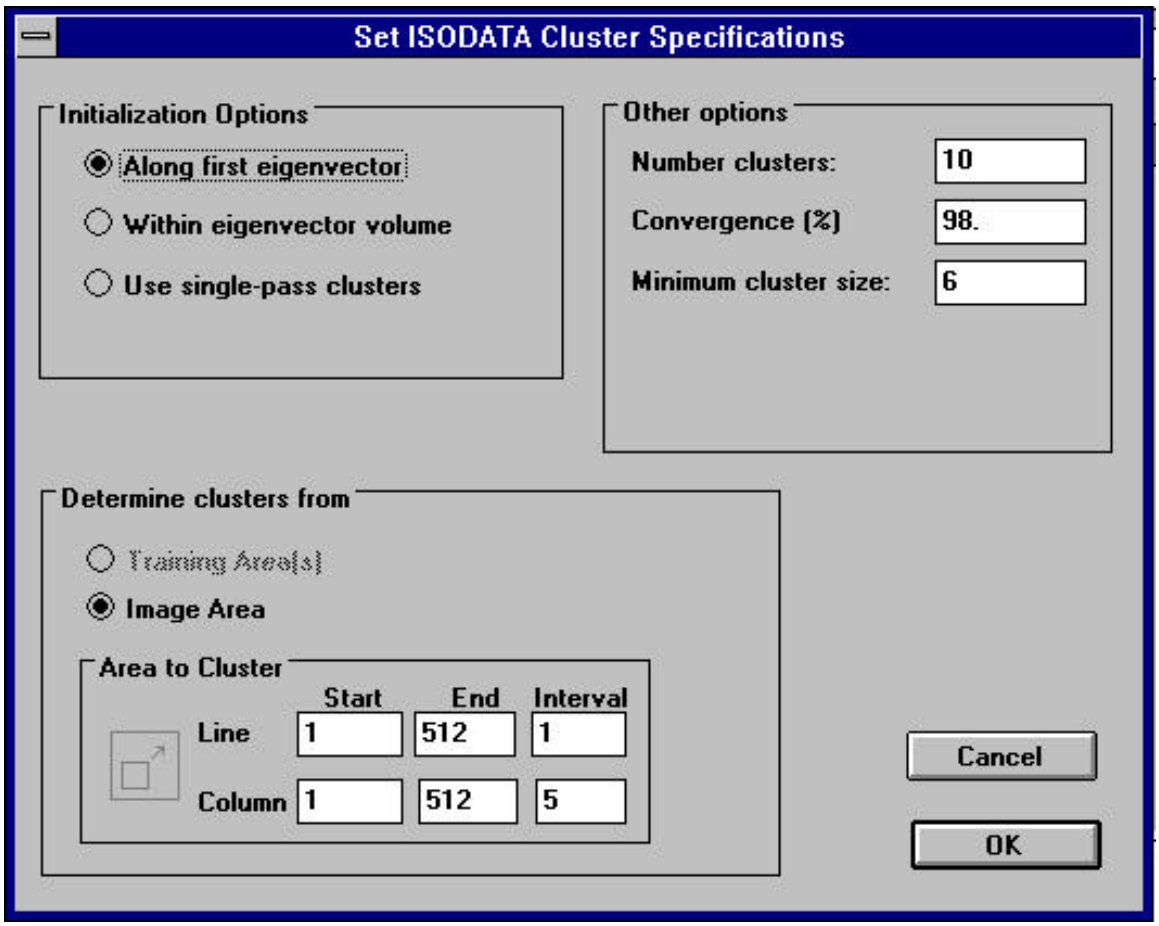


• First, click the **Image Area** button to place a dot in it.

• Click to place a marker in the **Disk File** box. This saves your project to disk.

• Lastly, click the **ISODATA** button, as indicated by the cursor in the diagram above. ISODATA is the algorithm, or mathematical process, that MultiSpec© will use in the clustering process.

A new window, the **Set ISODATA Cluster Specifications** window will open, as shown below.



It is in this window that you tell MultiSpec© how you want the clustering to proceed. The information you need to provide is:

• Be certain that the **Image Area** button is checked, as shown above.

• Select "**Along first eigenvector**." This is the specific algorithm[1] that MultiSpec© will use in its clustering.

• Leave the settings in the **Other options** boxes unchanged for this exercise.

Notes: "**Number of clusters**" tells the software how many different groups you wish for the classification For this tutorial, we will use the default setting of 10. The number you will use when you cluster your 512 x 512 image will be discussed later.
During the classification, the program goes through the data over and over. This is called "iteration." Each iteration is called a "pass." The system makes "passes" through the image until a preset percentage of the pixels in the image are **not** changed during the pass. The clustering then ends. This percentage is called the "**Convergence**."

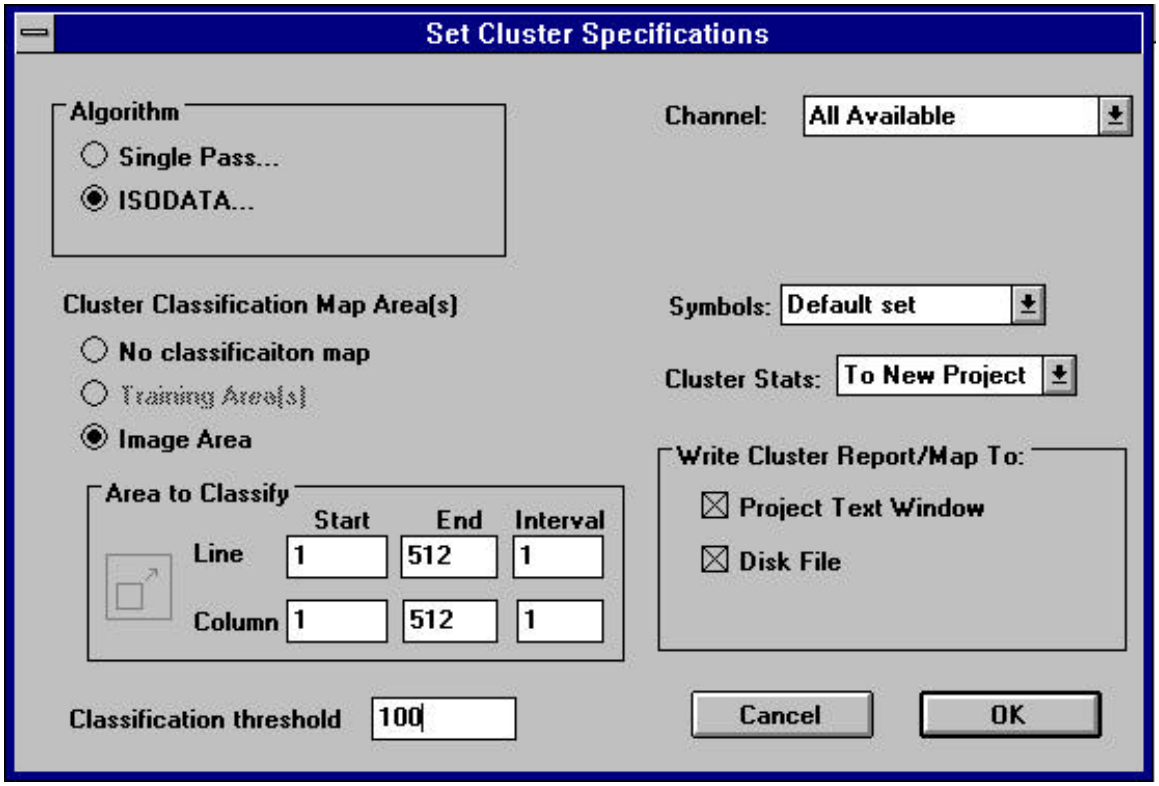
"**Minimum cluster size**" tells the system the smallest sized area to work with. Areas smaller than this minimum size will not be clustered.

1. For a discussion of MultiSpec©'s algorithms, see "An Introduction to MultiSpec©," by David Landgrebe and Larry Biehl, Purdue Research Foundation, 1995. This document may be downloaded from the Purdue/LARS WWW site at:

http://dynamo.ecn.purdue/~Biehl/MultiSpec©

• After you have made these settings, click **0K**

• The **Set Cluster Specifications** window appears again.



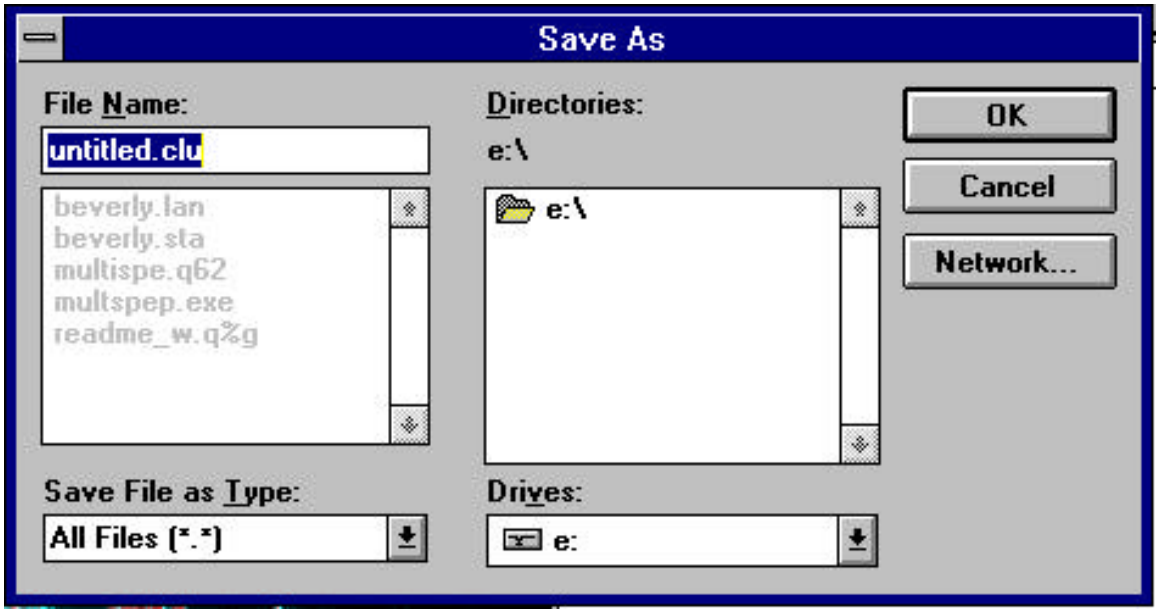
• In the lower left-hand corner of the box is the "**Classification threshold**" entry box. **Change the value in this box to "100"** just as you would change any item in a word-processor.

Setting this "threshold" value to 100 forces the system to assign every pixel in the image to one of the clusters. A value of less than 100 specifies the tolerance for assignment of pixels. A value of less than 100 will result in some pixels not being assigned to clusters.

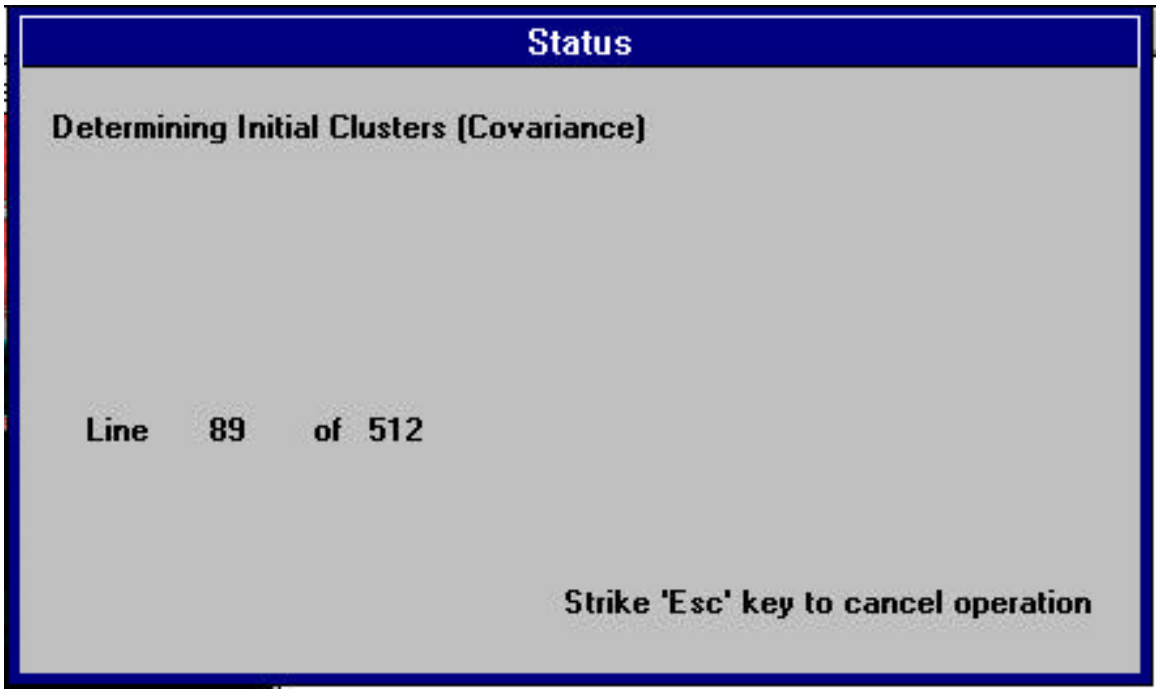
In this clustering, you are interested in large, fairly homogeneous areas, so individual pixels of slightly different spectral characteristics dotting the map are unnecessary.

• Click **0K**

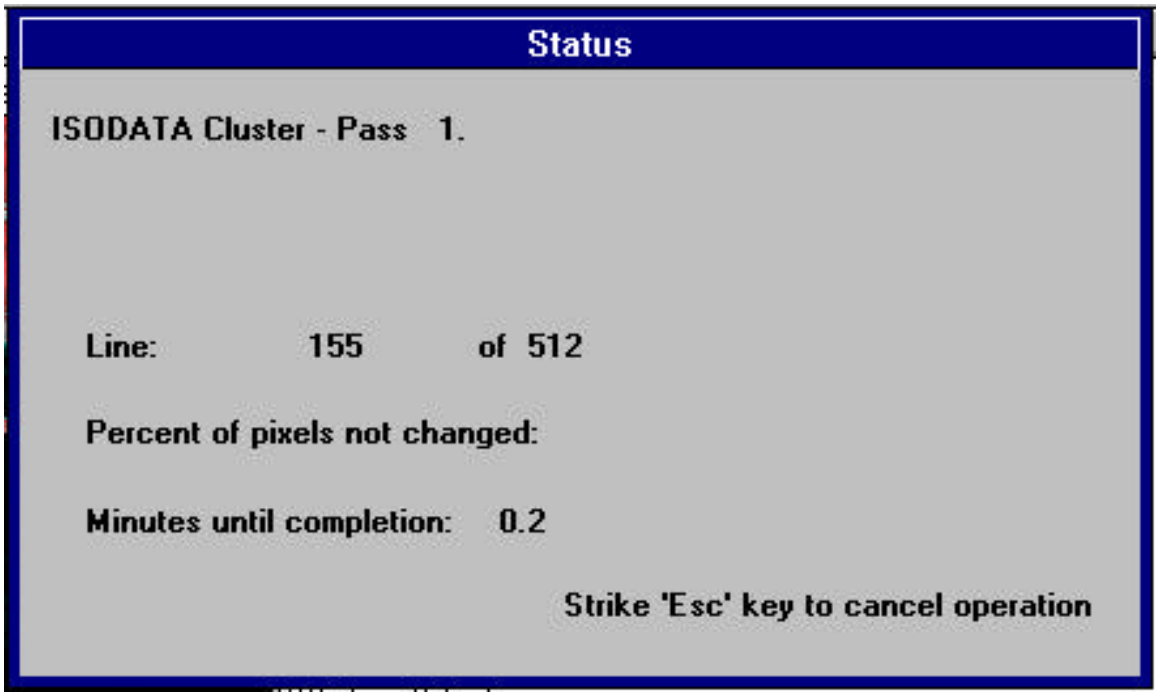
• The **Save As** dialog box appears, as shown below. There is a default name for your classified image file "**untitled.clu**." Leave the file name as Untitled.clu for this exercise. **Do not use the name Beverly.clu**, as that will be used for another file necessary for this tutorial.



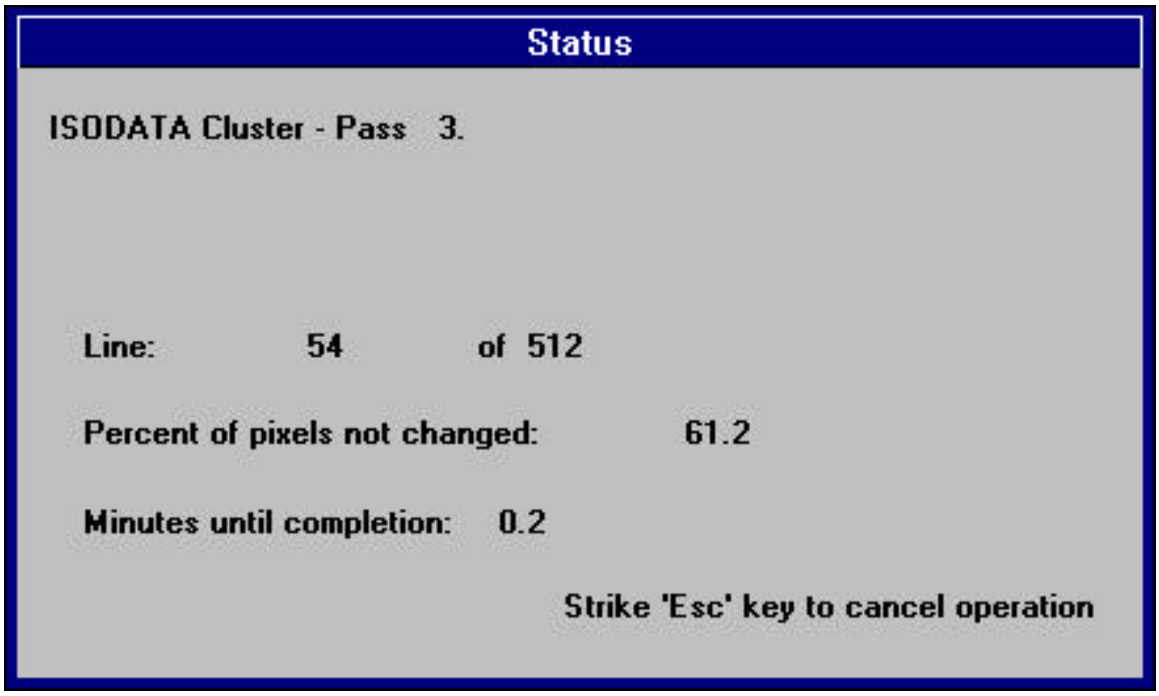
• The system then makes its first pass through the image to initially determine the clusters present as shown in the **Status** box, below.



• The "**Pass 1**" clustering **Status** box then appears, as shown below. During this initial iteration, Pass 1, the "Percent of Pixels Not Changed" shows no value. Also note that a time is given for completion of this operation.



• The **Percent of Pixels Not Changed** entry does not change until the end of Pass 2. During Pass 3, the value will be displayed, as shown below, along with a time to completion of the pass.

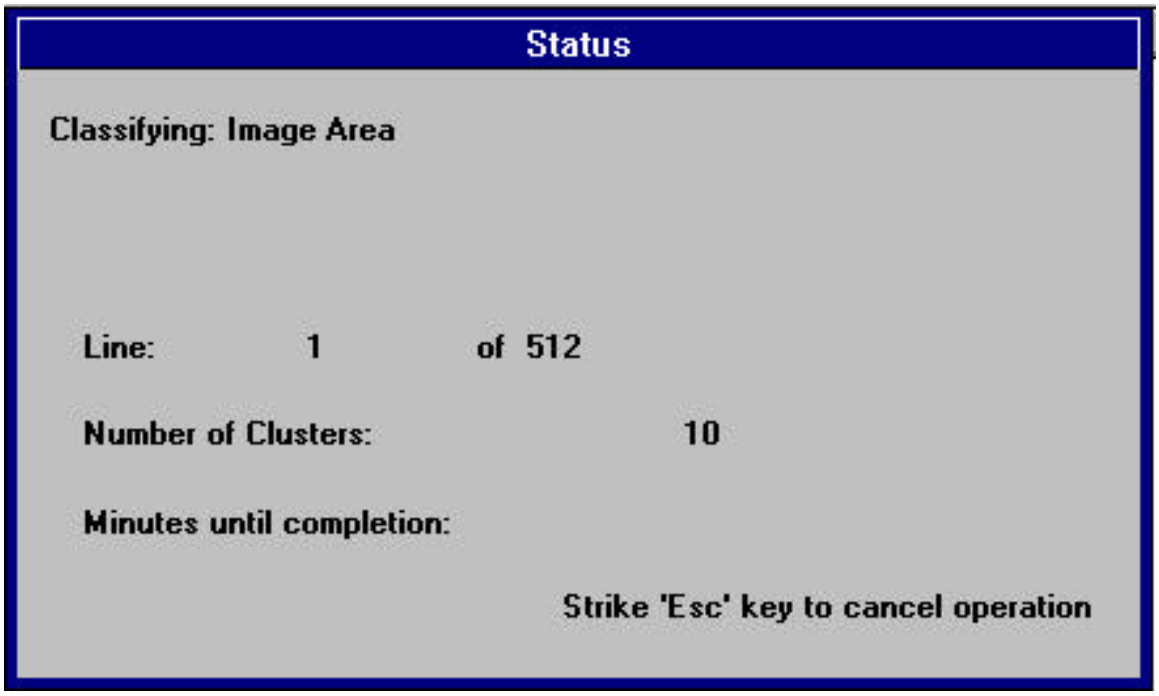


• During subsequent passes, the "**Percentage of Pixels Not Changed**" increases, until it reaches the value given in the "**Convergence (%) specification**." The time for each pass to be completed is given in this window.

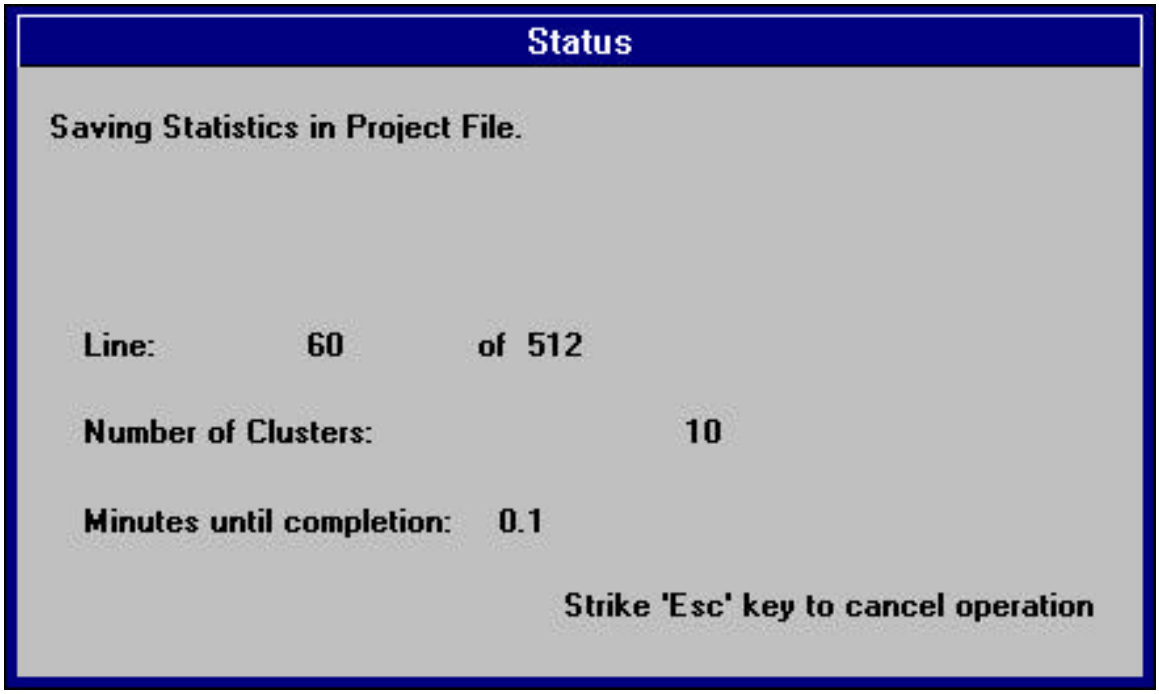
You can expect the system to make at least a dozen passes to achieve a 98% Convergence.
The time required for this process is dependent upon the processing speed of your computer. On a 386-based machine (the minimum processor requirement), you can expect the entire process to take several minutes. On a Pentium machine, the process is done very quickly.

Note: If you press "**Cancel**" during a pass, you will be asked if you wish to cancel immediately, or complete the iteration. Canceling immediately terminates the clustering, while finishing the iteration ends the clustering at a Convergence less than initially specified.

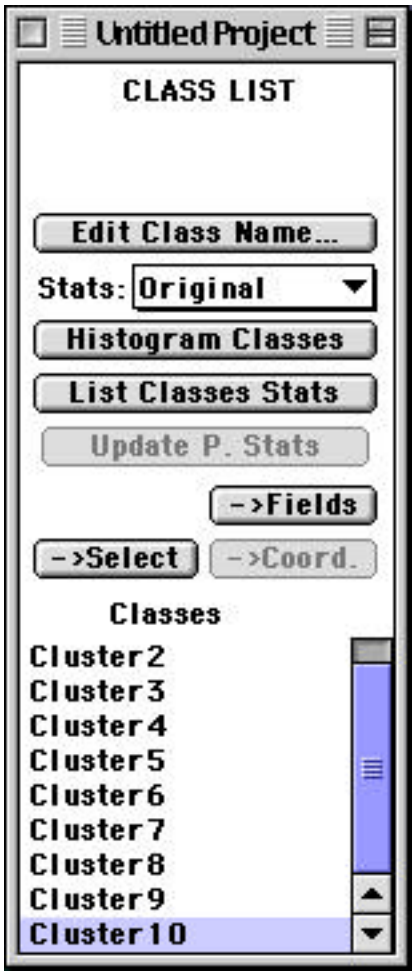
• After the clusters are determined, the system will display the "**Classifying Selected Image Area**" window, below. Here the system assigns individual image pixels to the clusters it has determined.



• After the clustering is complete, you will see the **Saving Statistics in Project Fi l e** window, shown below.



- After the clustering process ends, the **Untitled Project Window** appears, as shown to the below.



- **Close** this window.

- When asked is you wish to "**Save 'Untitled Project'** before closing?", click **Don't Save.**

It is not necessary to create or save a project for the Unsupervised Clustering. For this work, we do not need to access the statistical information that is saved with a project.

The Results of Clustering

There are two results of clustering:

A **description of clustering activity** and a “text map” in the **TEXT OUTPUT** window,

A **clustered Thematic image**.

• From the **Windows** menu, select **untitled project**. Scroll to the top of this “text window," and you will have statistics describing the clustering and its results. A part of the text output for the sample clustering is shown below. In it are listed the number of clusters produced and the average value (mean) of the pixel values for each band in each of the classes.

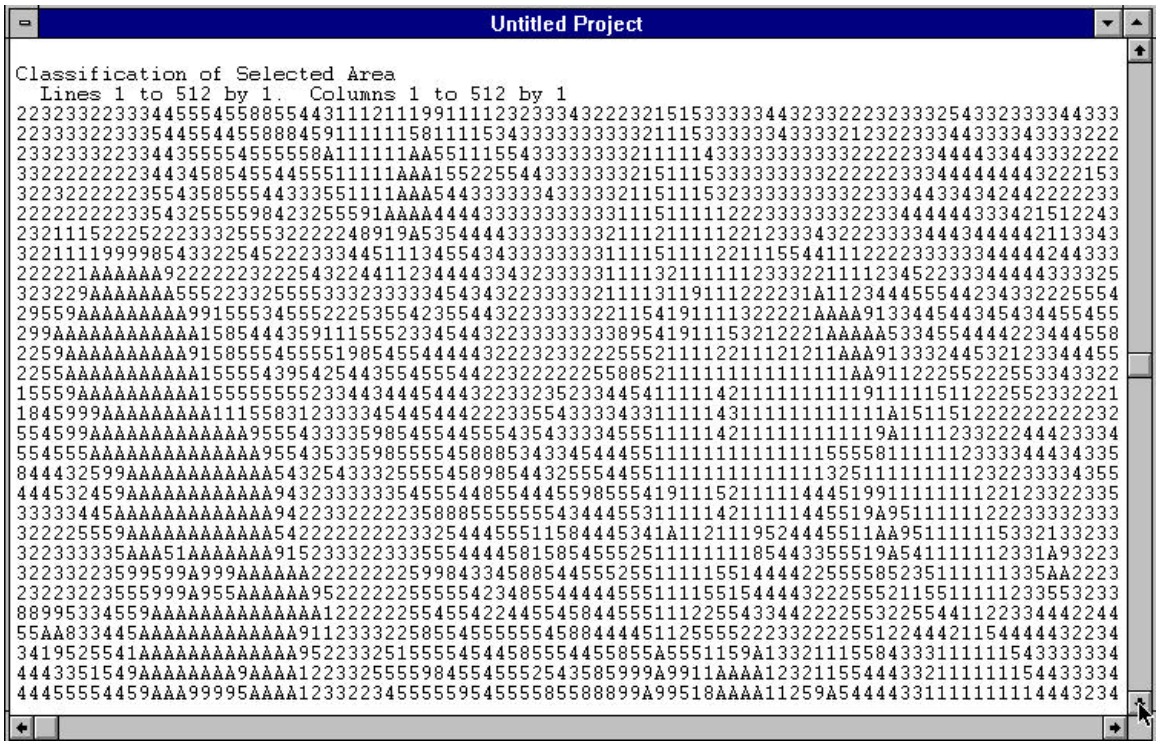
```
Clustering completed after 6 passes and 929 of 52429 pixels changed.
Final cluster class statistics.
Cluster  Pixels  Channel Means
                    1      2      3      4      5
    1     1847   114.6 138.5 114.8 216.8 220.0
    2     6031    69.5  85.5  59.4 221.3 160.5
    3     6558    59.9  71.1  48.3 188.8 133.4
    4     3688    75.4  78.2  63.6 144.6 117.1
    5     3761   120.9 120.9 106.9 160.6 151.4
    6    22690    58.3  39.2  23.1  12.8   8.7
    7     2345    96.8  79.9  55.7  22.7  19.4
    8     2372   163.4 147.6 139.1 116.2 127.3
    9     1894   205.2 188.8 182.3 128.0 154.1
   10     1243   248.5 245.6 243.3 134.4 184.6
```

Also produced is a text map of the clustered area. The system assigns a number or letter to each of the clusters, and then displays a map of the clustered area using this code. For the clustered Beverly.lan image, the code is shown below.

```
Classes used:
  0: Thresholded
  1: Cluster 1                          1
  2: Cluster 2                          2
  3: Cluster 3                          3
  4: Cluster 4                          4
  5: Cluster 5                          5
  6: Cluster 6                          6
  7: Cluster 7                          7
  8: Cluster 8                          8
  9: Cluster 9                          9
 10: Cluster 10                         A
```

A portion of the Text Map from the Clustering process appears below. Each number/letter represents a pixel and the clustered group to which it belongs. You can see,

even from this representation, that the system has identified several large, homogeneous areas, identified by the appearance of the same letter/number in an area.



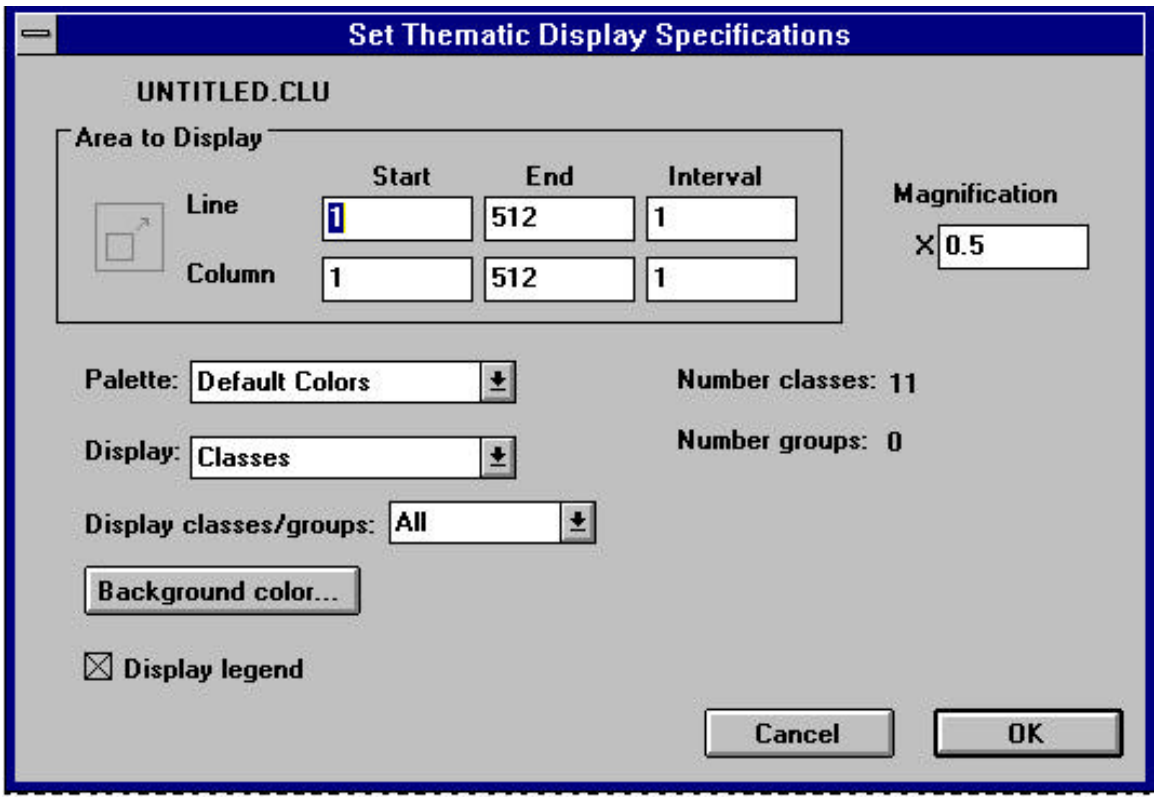
Examining the Clustered Image

• From the **File** menu, select **Open Image**.

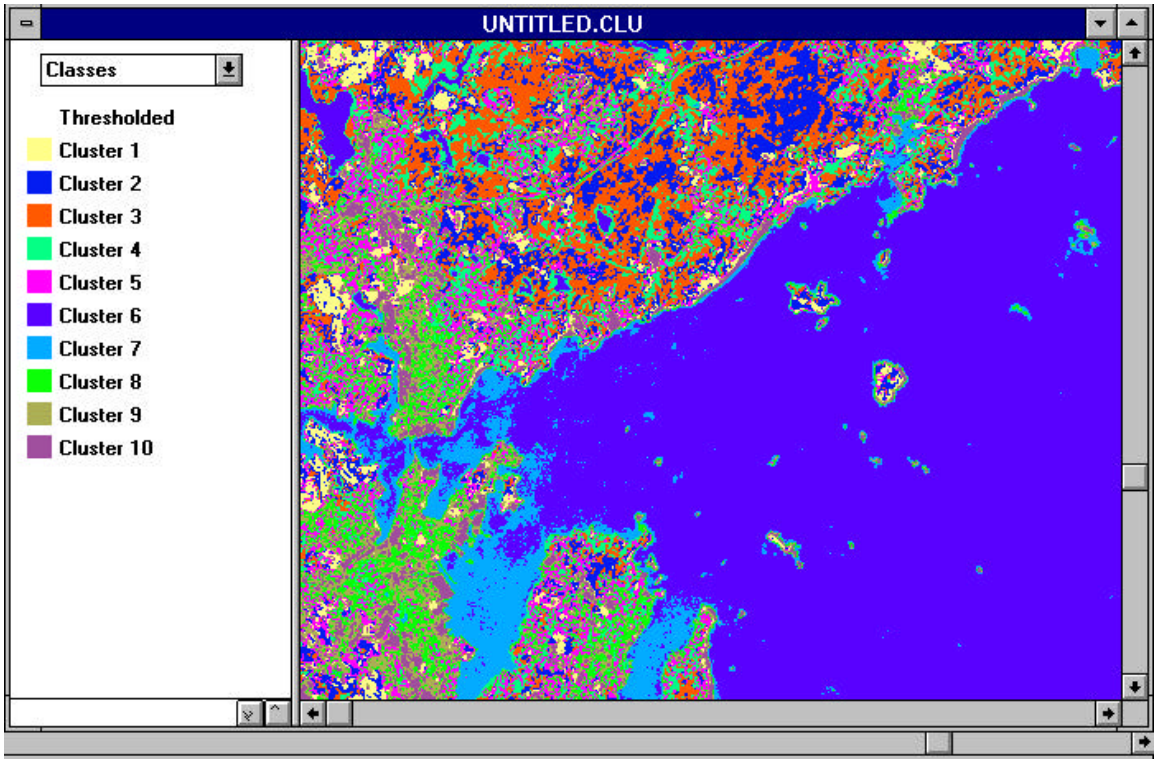
• Select the **Untitled.clu** file name you used earlier, and click **Open**.

• The **Set Thematic Display Specifications** window opens, as shown below.
You can experiment later with some of the other palettes in this menu, but for now accept the default settings and press **0K**.

You will note that this window lists **11** classes. One of these is called the "**Thresholded**" class, and would normally contain pixels not included in the clustering process.
However, recall that you set the Threshold at 100, so this will be an "empty" class.



• Your clustered image opens as shown below.



• Notice that there are 10 numbered classes, plus a class labeled “Thresholded.” This “Thresholded” class contains no pixels, because you set the “Thresholding” value to 100, on page 6. Each class is assigned a color by the system which has nothing whatsoever to do with what the cluster represents. The clusters are produced and arranged in order of descending level of brightness. That is, clusters near the top of the list represent surface materials that are “brighter’ (have greater reflectance) than those near the bottom of the list.

• You may print the image from the **File** menu. When you do, the clustering key will be printed along with the image.

• You may use some of MultiSpec©’s regular tools with this Thematic Map. Such tools as: the **Zoom** feature, and the **Coordinate Bar**, from the **View** menu, function normally. The **New Selection Graph** feature will show a plot with only one piece of data. This map is no longer “multispectral.” Each pixel no longer contains data for different LandSat bands, or channels. Each pixel contains only one value, which identifies its cluster.

• The software now incorporates the ability to change the color of each cluster’s “color chip” to one appropriate to your scheme of land cover classification. This process will be detailed later in this tutorial.

• If you do a clustering with a larger number of classes, you may not be able to see them all in the “Classes” column. To scroll through this column:

- Move your cursor into the column

- Hold the mouse button down
- Drag to either the top or bottom of the column.

The classes will scroll up and down.

• You and your students will be preparing a thematic map from this clustered image in which you identify some of the clustered areas by their actual land cover. To do this, you may save the image as a TIFF file from the File menu. This process does not save the clustering key, only the image area will be saved. The TIFF file may then be brought into any one of a number of paint or draw programs to be “fancied up” as a thematic map.

• If you wish to have an image that contains the clustering key, and can also be moved into paint or draw programs you can capture the entire screen using one of a variety of “screen capture” programs that are available in the public domain or as “shareware.” You will want to examine the features of these to determine that they save “captures” in a format that can be read by your paint/draw program.

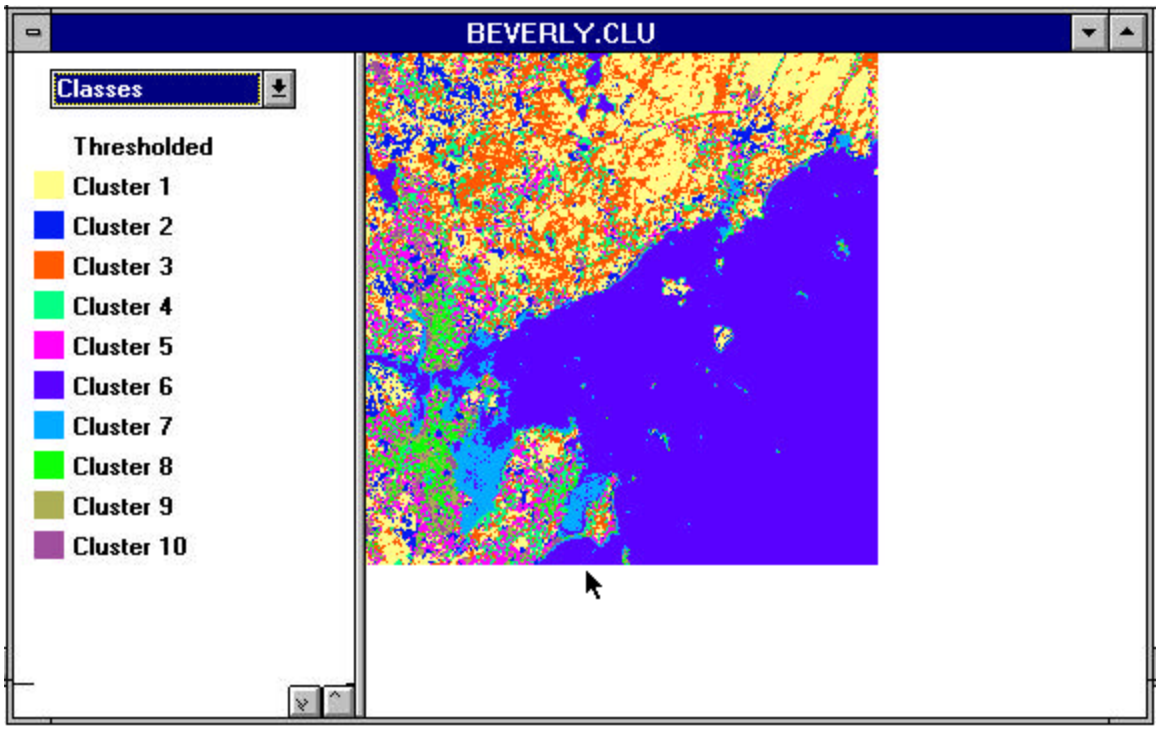
How Valid Is the Clustering Process?

It is necessary for you to be confident that this process of “unsupervised classification” actually yields clusters that are related to land cover types. To this end, included with this tutorial is a file named “**beverly.clu**." In your **Unsupervised Clustering**, you had the computer produce "clusters" based on similarities in spectral characteristics among pixels. In a **Supervised Classification**, the operator, who is familiar with the land cover in the region, first "trains" the computer to recognize different land cover types. The computer than classifies the image based on these known land cover types. The **Beverly.clu** image contains this kind of information.
.

• Be certain your “**Untitled.clu**” image is open.

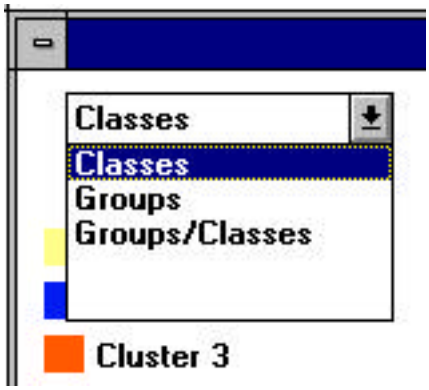
• From the **File**, select **Open**.

- Open the **Beverly.clu** image.
- At the **Set Thematic Display Specifications** screen, press **0K**.
- The image will open as shown below.

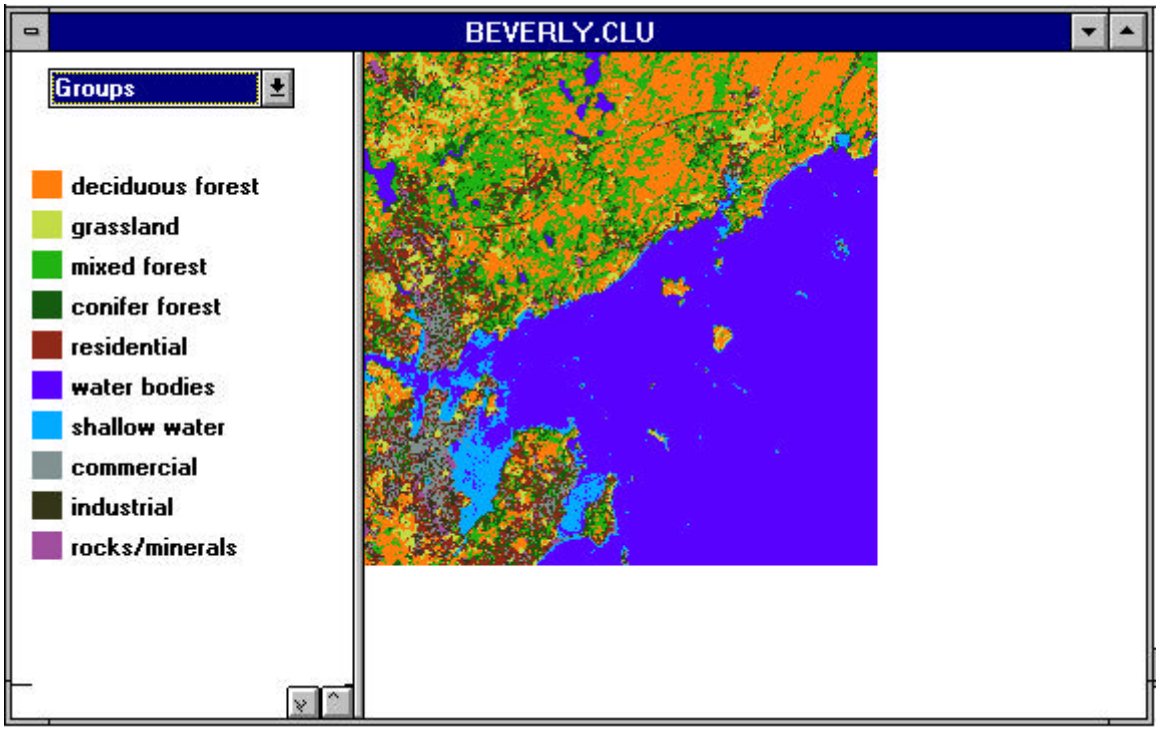


This image looks very much like your own clustered image, with no land cover identification. However, in addition to the clusters (which MultiSpec© calls “classes,”), it also contains “group” information, which is the results of the supervised classification.

- From the **Classes** menu, as shown below, select **Groups**.



The legend now displays information from the supervised classification.



• Resize the images and arrange them side-by-side on the screen. (See Appendix 1 for directions.)

• Compare the areas identified in the supervised classification (**the Beverly.clu** image) to the clusters produced by the system in your unsupervised clustering (the **Untitled.clu** image.)

You should see that the unsupervised clustering provides, at least in this case, a good indication of the locations of large areas of uniform land cover that could be investigated for verification studies.

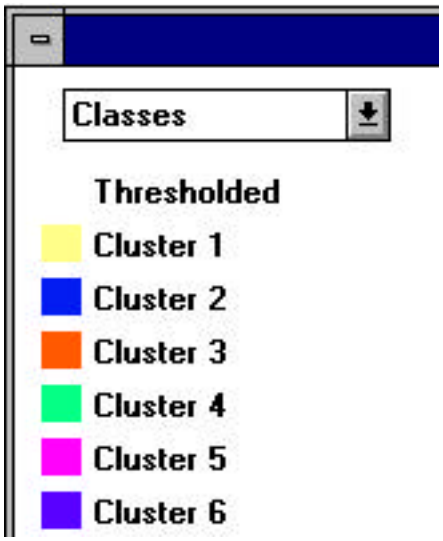
Renaming the Clusters

• **Close** the **Beverly.clu** image by selecting **Close Window** from the **File** Menu.

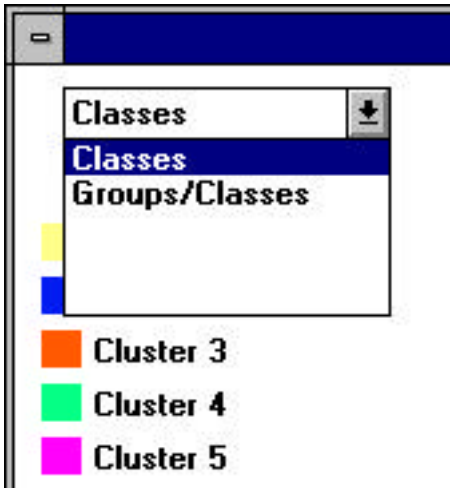
Your unsupervised clustering produced clusters identified only by a number, and arranged in order of decreasing brightness. Once you have identified the land cover for each of these clusters, your Thematic Map display may be customized to show these clusters either by name or by **MUC** identification code. You can, in effect, produce two different Thematic Maps on the same image; one in which each cluster is identified by a name (e.g. Ocean, Transportation) and the other by **MUC** designations (e.g. 72, 93.)

The secret to this process is that your Thematic Map can display both "Groups" and "Classes." When it is produced, both "Groups" and "Classes" have the same set of colors and labels. To see this:

• Click on the **Classes** pull-down menu shown in the diagram below.



• The pull-down menu will show the choices illustrated below.



• Select "**Groups/Classes**." Then immediately pull down the window again and select "**Groups**."

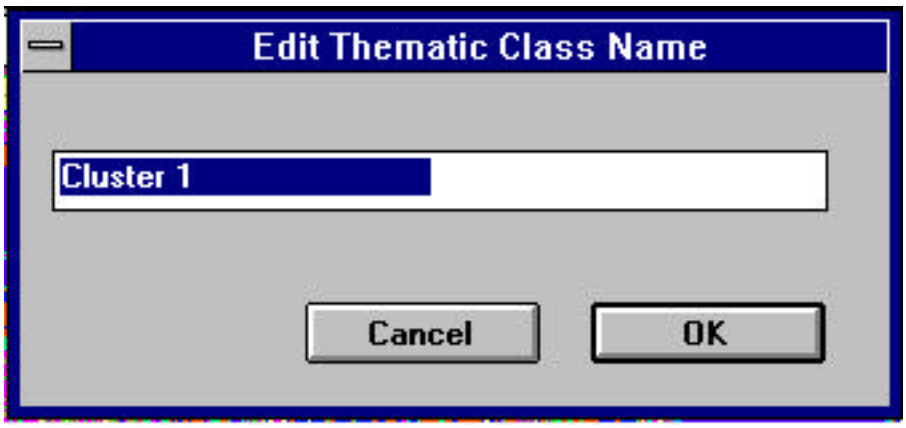
• You can now switch between "Groups" and "classes" and, for now, you will see for now that the information in each view is identical.

You might decide that the "Groups" will contain descriptive names, while "Classes" contains MUC labels.

To change the name of a cluster in either view, at any time:

• **Double Click** on the cluster name (for example, "Cluster 1")

• The **Edit Thematic Class Name** dialog box, as shown below, opens.



• You may now enter either a descriptive name or MUC identification number for this class.

• Once you enter a descriptive name in, say, “Groups,” use the pull-down menu to select “Classes” and enter the corresponding MUC identification number for that same Cluster.

The October 1998 release of MultiSpec© PC will save both cluster and class information, while older versions do not. If you are not using this version, you should visit the MultiSpec© page at Purdue University to download it.

• From the **File** menu, select **Save Thematic Group Info**.

• From the **Project** menu, select **Save**

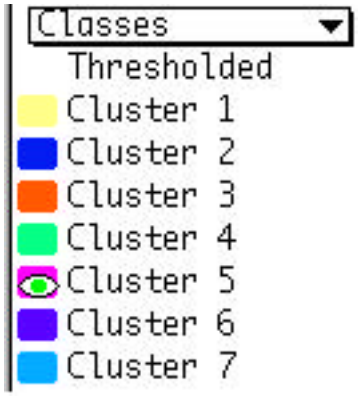
.

• Accept the default name, and click **0K**

Changing Cluster Colors

It is sometimes difficult to tell which color "chip" in the Classes column is associated with a colored area in the image. To match classes to their image areas:

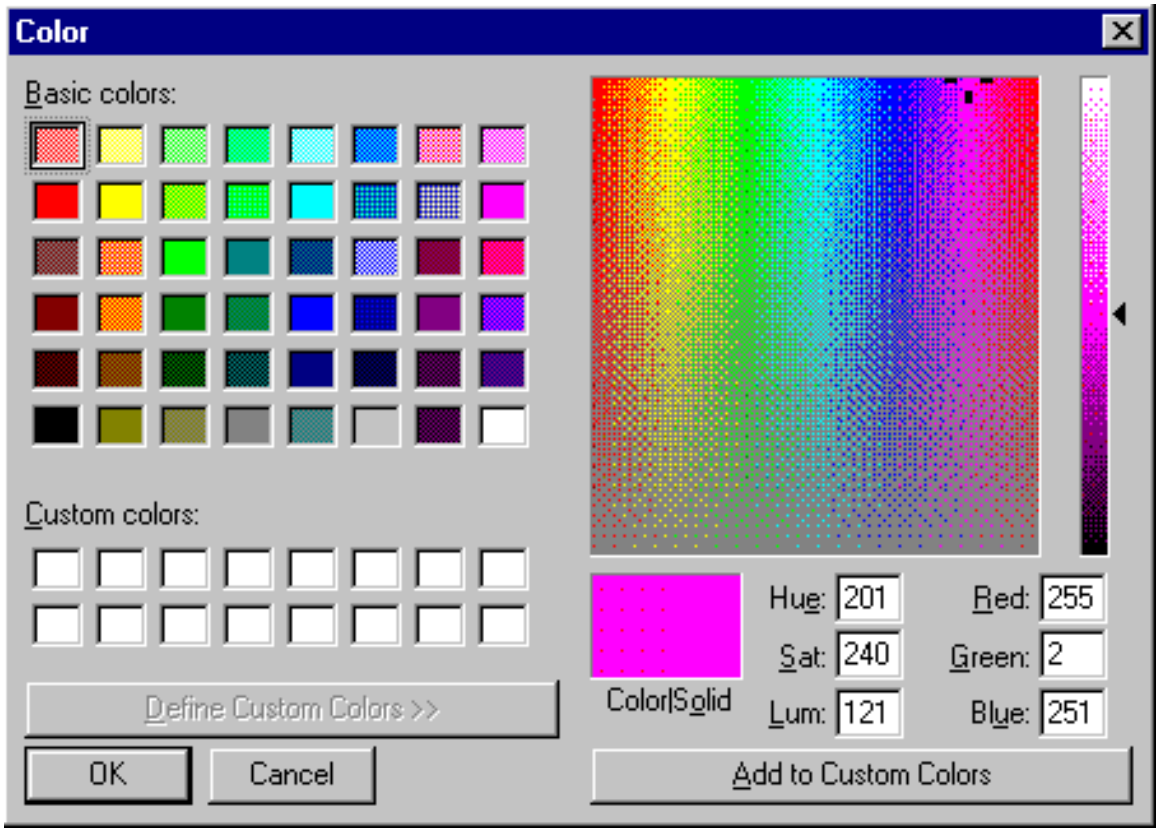
- Place the cursor over any color box in the "Classes" column.
- Hold down the **Shift** key: The cursor changes to an "eye," as shown below over cluster 5:



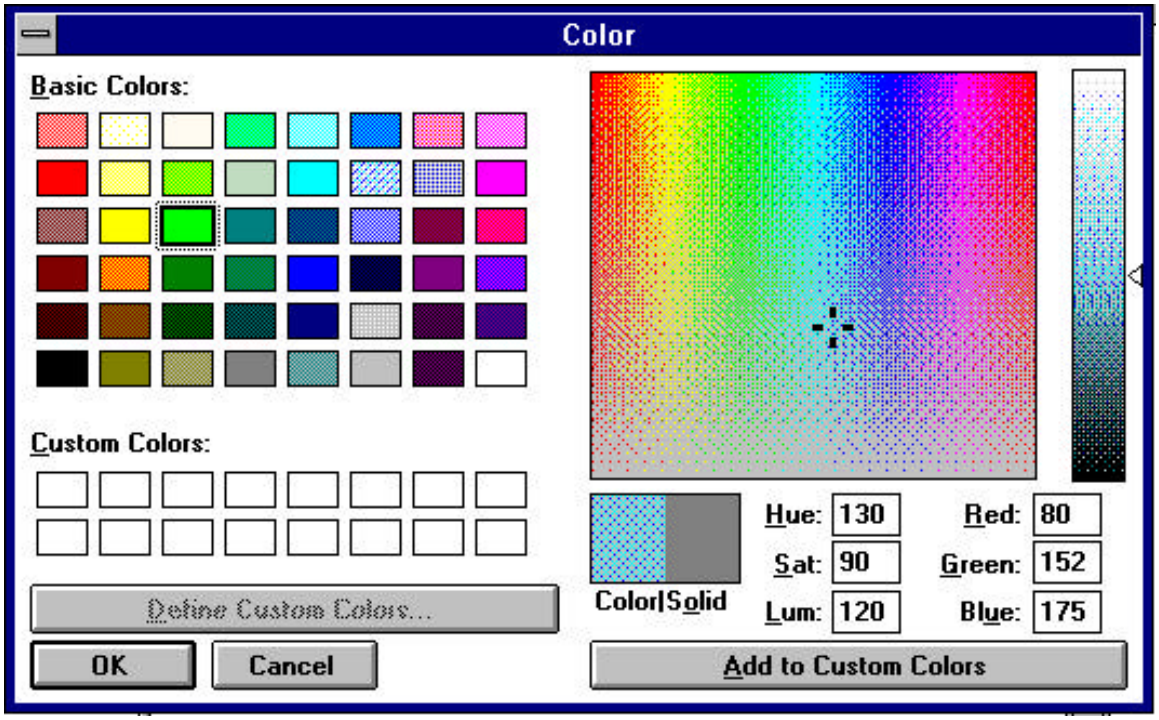
- While holding the shift key down, **Depress the mouse button** and the areas in the image of that class will "blink," or turn to white.

To Change the Color of a Cluster:

- Double click on any cluster color chip.
- The "Color" window opens, as shown below.



• The currently selected color is shown in the "**Color/Solid**" box, and its position is indicated by a "cross" in the upper right-hand corner of the color chart.
• You may select any one of the "**Basic colors**" by clicking once on its color chip, and then clicking **0K**.
• You may define a custom color by clicking anywhere in the color chart. A "cross" will mark your selection, as shown below.



• Your selected color will also appear as a chip in the "**Custom colors**" area.

• You can change the intensity of the selected color by adjusting the "slider" to the right of the color box.

• After you have selected a color for a cluster, click **0K** and the cluster and its chip will change to the new color

How Many Clusters Do I Use

Most regions the size of your 15 km x 15 km GLOBE Study Site do not generally contain a large number of different land cover types. When you first perform a clustering on your 512 x 512 image, use the same values as you used in this tutorial. Examine the results in light of your knowledge of your own area. Do some fieldwork and look at the areas your clustering suggests are fairly large and homogeneous. Compare your findings to the MUC Classification System. Only if you feel that this clustering does not adequately represent the land covers in your area should you increase the number of clusters, and then 12 to 14 clusters should be sufficient to do the job.

Reporting the Data

In order to report your data, you must make some "sense" out of the clusters determined by this unsupervised process. You can then relabel the clusters as what type of land cover they represent. The process involves the following steps:

- Desk Verification
- Field Verification
- Completing the Accuracy Assessment for your land cover map *
- Renaming the Clusters
- Sending in your completed map.

(* Before you can submit your map to GLOBE, it is necessary for you and your students to determine how accurately you identified your land cover elements. See the **Accuracy Assessment** section of the Land Cover /Biometry module of your GLOBE Teachers Guide.)

Desk Verification

This process involves your use of local maps (topographic, land cover, soil, political, etc.), other local references (aerial photos, people, agencies, etc.) and the combined experiences of both you and your students to identify some of the clusters produced by MultiSpec©. Use whatever resources you can to identify them. Remember that your identifications should correspond to the most detailed level of the MUC (Modified UNESCO Classification System available for that type of land cover.

Field Verification

If there are clusters that you cannot identify "from the desk," you will have to go out into the field to determine what they are. If a formal "field trip" is not possible, in all probability someone lives near to or drives by that area and can do the identification.

A Note About Expectations, and a Caveat

When you proceed to the classification of your own 512 x 512 image, you will find the appearance of your clustered image probably considerably different than this demonstration. Certainly the nature, abundance and distribution of land cover types in your image will differ from those in the Beverly, Massachusetts area.

As you cluster your own image, you may find that specifying 10 clusters does not discriminate between standing bodies of water, except perhaps between fresh and salt water. In other words, lakes, ponds, rivers, etc. will probably all be clustered into the same group, unless there are significant surface properties that might change their reflectance (i.e., significant algal growth on the surface of a farm pond.)

Validating Your Results

Before submitting your classified land cover map, be certain to the Accuracy Assessment Protocol in the Land Cover/Biology section of your GLOBE Teacher's Guide.

Submitting your results:

Once you have an unsupervised classification (clustering) that seems to adequately represent your 15 x 15 GLOBE Study Site, your results will be submitted to the GLOBE Map Archive, as described in your Teachers Guide.

• Make a copy of your clustered Thematic image onto a high-density floppy disk and clearly label it with your school name, your name, and "clustered image."

• Using your favorite word processor, prepare a file with the following "metadata:"

Your School Name
Your Name
School Address
Date your image was acquired*
The LandSat "path and row" data for your image.*
Some information about yourself, your students, and some of your experiences in doing your clustering.

*This data should be printed on the color prints of your GLOBE Study Site provided to you by the GLOBE Program. From your word processors options, save this data as a text file and place it on the same disk with your completed land cover map.

• From your word processors options, save this data as a text file (or ASCII file) and place it on the same disk with your completed Land cover map.

• Carefully package these materials and send them to:

GLOBE Student Data Archive
NOAA/NGDC E/CG 1
325 Broadway
Boulder, CO, USA 80303

Appendix 1:

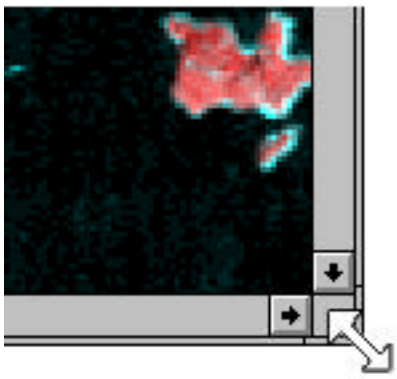
Viewing Two Images Side-by-Side in MultiSpec©: PC

In MultiSpec©, you may simultaneously view two (or more) images or view the same image in more than one band combination.

To view two images side-by-side:

• **Open** the first image in whatever band combination you wish.

• Place your cursor over the lower right-hand corner of the image, as shown below. It will turn to a double-headed arrow (size exagerated.)



• Hold your mouse button down and **drag** the image size to one-half your screen size.

• From the **File** menu, select **Open Image**.

• Select the next image and band combination you desire. You may open a different image, or the same image in a new band combination. This allows you to directly compare the differences seen in two or more band combination.

• Again click on the lower-right image corner and resize this image to one-half screen.

• Click on the second image's gray **Title Bar**, at the top of the image, and drag the image to the right-hand half of the screen.

• Within the constraints of your screen size and available memory, you may open as many image windows as you wish. These may be different images, or the same image in different band combinations.